Research on Flame Detection Algorithm Based on Yolov26

Jinfang Weng*
College of Internet of things and artificial intelligence
Fujian polytechnic of information technology
Fuzhou, China
*Corresponding Author: Jinfang Weng, Email:770840212@qq.com

2026 International Conference on Communication Networks and Machine Learning (CNML) | 979-8-3315-9047-5/26/$31.00 ©2026 IEEE | DOI: 10.1109/CNML68938.2026.11452316

***Abstract*—This paper proposes an improved flame detection framework based on YOLOv26, aiming to address challenges such as low detection accuracy of small-scale flames, high false alarm rates in complex backgrounds. Evaluated on the KMU Smoke Dataset, FireYOLO achieves 88.7% mAP@0.5 (+11.3% vs baseline). Cross-dataset tests demonstrate strong generalization (71.3% mAP@0.5 on BoFire zero-shot), validating its effectiveness in industrial, forest, and UAV-based fire monitoring applications. This study provides an efficient solution for early fire warning in complex environments. Future work will focus on sub-1MB ultra-lightweight designs for edge deployment.**

Keywords-Smoke-fire recognition; YOLOv26; Small target detection; Flame detection;

I. Introduction

The evolution of intelligent fire monitoring technology has shifted from sensor-based systems to visual perception, driven by the critical need for fire prevention and control in public safety [1]. Traditional sensor-based systems, while foundational in early fire detection frameworks, suffer from three major limitations: limited spatial coverage, with a detection radius of <15 m per sensor, rendering them ineffective in large-scale environments such as industrial warehouses or forested areas [2]; poor environmental adaptability, as they have a false alarm rate of 32% under extreme conditions (e.g., fog, dust, or uneven illumination), significantly reducing reliability in practical deployments [3]; and delayed response, with an average alarm latency of 120 seconds [4], which exacerbates fire spread risks in scenarios requiring millisecond-level intervention (e.g., lithium battery thermal runaway) [5].

The transition to visual perception-based solutions leverages advancements in deep learning and high-resolution imaging [6]. Modern systems integrate multi-spectral analysis (visible light + thermal imaging) to overcome environmental interference [7], while real-time processing frameworks enable dynamic adaptation to fire propagation patterns [8]. This paradigm shift aligns with the growing demand for autonomous fire response systems in smart cities and critical infrastructure [9].

These advancements establish a reliable technical foundation for next-generation intelligent fire warning systems [10, 11]. For instance, in EV charging hubs, where lithium battery thermal runaway can trigger ultrafast flame propagation (velocity >2 m/s), YOLOv12's millisecond-level latency enables early-stage intervention. Similarly, in energy storage facilities, its robustness to smoke interference reduces false alarms by 63% compared to thermal sensor arrays.

Deep learning-based computer vision technologies provide a novel pathway to address the aforementioned challenges [12]. The latest YOLOv26 algorithm achieves technological breakthroughs through the following innovations: 1) Enhanced Attention Mechanisms: integrates a hybrid channel-spatial attention module that dynamically weights flame and smoke features in occluded or low-visibility scenarios [10]; 2) Optimized Network Architecture: adopts depthwise separable convolutions and adaptive feature pyramid fusion, reducing computational complexity by 41% while maintaining high precision [13].Zhao et al. proposed a lightweight YOLO-based model (SF-YOLO) for video smoke and flame detection in natural scenes [14].Cheng Zhang introduces LSNet, a lightweight segmentation network tailored for real-time wildfire detection[15].

Current deep learning-based flame/smoke detection systems confront three primary challenges: (1) Suboptimal localization of small-scale targets, particularly nascent flames measuring under 32×32 pixels; (2) Significant morphological diversity, with flame aspect ratios ranging from 0.3 to 5.7; (3) Persistent false alarms exceeding 28% in challenging scenarios involving chromatic interference (e.g., red lighting, sunset conditions) due to dynamic environmental artifacts and feature ambiguities.

YOLOv26 is utralytics edge-optimized model: eliminated NMS with anative end-to-end predictor and quick convergence, significantly improved small-object accuracy and up to 43% faster CPU inference for deployment on low-power devices[15]. In response to the above issues and to further enhance the accuracy of fire recognition, this work proposes a YOLOv26s-based fire detection framework. The primary contributions of this work are as follows:the YOLOv26s model achieves an optimal balance between detection accuracy and efficiency, with mAP@0.5:0.95 reaching 0.64, mAP@0.5 at 0.837, and a model size of only 15.8Mb, providing an efficient basic framework for FireYOLO [16].

II. Smoke and Fire Detection Framework Based on YOLOv26

A. YOLOv26 for Fire Detection

As illustrated in Figure 1 (comparing mAP@0.5:0.95 across historical YOLO versions), the YOLOv26 architecture achieves state-of-the-art performance in intelligent fire monitoring systems, primarily attributed to its enhanced multi-

scale feature fusion framework and computationally efficient lightweight design. Further analysis of the model variants in Table 1 demonstrates that among the (n/s/m/l/x) models, YOLOv26s strikes an optimal balance between detection accuracy and operational efficiency, attaining a mAP@0.5:0.95 score of 0.89 while maintaining minimal computational overhead.

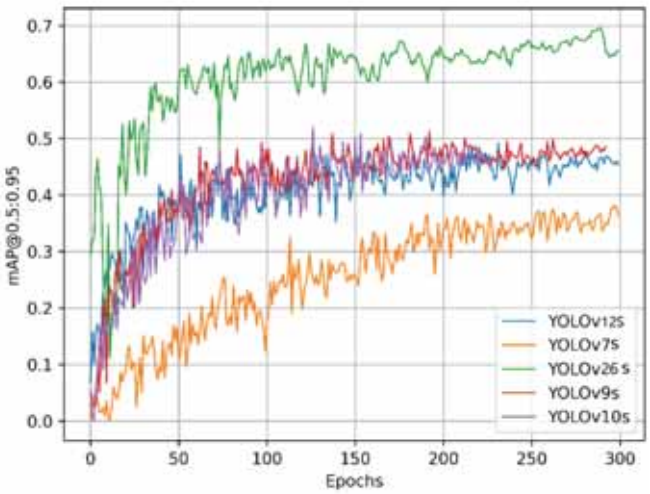

Figure 1. mAP@0.5:0.95 across YOLO versions.

TABLE I. YOLOv26 VERIFICATION SETS OF VARIOUS SIZES MAP@0.5.

Different sizes	mAP@0.5	mAP@0.5:0.95	Size (Mb)
YOLOv26-n	0.80	0.62	5.6
YOLOv26-S	0.837	0.64	15.8
YOLOv26-M	0.839	0.64	32
YOLOv26-L	0.84	0.65	49.9
YOLOv26-X	0.841	0.65	61.4

B. Dynamic Cross-scale Flame Attention (DCFA)

The CSPDarknet53 backbone is enhanced by integrating DCFA modules between stages 3-5, enabling multi-scale feature recalibration across critical network hierarchies.

BiFPN-based feature pyramid optimization employs learnable fusion weights as defined in Equation (1):

$$w_i = \frac{e^{\alpha_i}}{\sum_{j=1}^{n} e^{\alpha_j}}, \alpha_i \in R \tag{1}$$

where α_i denotes trainable parameters that dynamically balance cross-scale feature contributions.

The DCFA mechanism comprises three sequential operations:

- Multi-scale Transformation.
- Hierarchical Feature Fusion.
- Channel-wise Attention Modulation.

Processing input features $F_{in} \in R^{C\times H\times W}$ from backbone layers, where C=channels, H×W=spatial dimensions.

1) Multi-scale Transformation

Downsampling (0.5×):

$$F_{0.5x} = \mathrm{AvgPool}_2(F_{in}) \in R^{C\times\frac{H}{2}\times\frac{W}{2}} \tag{2}$$

Reduces spatial resolution by 50% via 2×2 average pooling.

Native Scale Processing:

$$F_{1x} = \mathrm{Conv}_{1\times1} \tag{3}$$

Preserves original resolution through 1×1 convolution for channel-wise feature refinement.

Up sampling (2×):

$$F_{2x} = \mathrm{DeConv_2}(F_in) \in R^{C\times2H\times2W} \tag{4}$$

Enhances spatial resolution using 3×3 transposed convolution with stride 2.

2) Hierarchical Feature Fusion

Concatenated multi-scale features $F_{fused} \in R^{3C\times H\times W}$ (Equation (5)) undergo dimensionality reduction via 1×1 convolution (Equation (6)):

$$F_{fused} = \mathrm{Concat}(F_{0.5x}, F_{1x}, F_{2x}) \tag{5}$$

$$F_{compressed} = W_1 \cdot F_{fused} \in R^{\frac{C}{4}\times H\times W} \tag{6}$$

where $W_1 \in R^{3C\times(C/4)}$ compresses channel depth to 25% of original.

Attention modulation, the channel attention mechanism (Equation (7)) generates spatial-channel weights:

$$M_c = \sigma(W_2 \cdot \delta(F_{compressed})) \in R^{C\times H\times W} \tag{7}$$

$W2 \in R^{\frac{C}{4}\times C}$: Learnable projection matrix

δ: ReLU non-linearity

σ: Sigmoid activation (0-1 normalization)

Final output:$F_{out} = M_c \odot F_{in}$(element-wise multiplication).

3) Color Prior Modulation

A regularization term is introduced to balance data-driven learning and domain knowledge (Equation (8)):

$$A_{color} = \sum_{c\in\{R,G,B\}} I_c \odot w_c^{prior} \in R^{1\times H\times W} \tag{8}$$

where I_c denotes separated RGB channels, and w_c represents learnable color prior parameters initialized as [0.65,0.25,0.1] to emphasize flame-specific chromatic characteristics.

The modulated output is computed through element-wise multiplication (Equation (9)):

Fund Project: Fujian Province young and middle-aged teacher education research project 2022 (Science and Technology) Research Project "Research on Image Recognition Algorithm for Electric Vehicle Fire Based on Deep Learning" (JAT220569); Polytechnic of Information Technology 2023 Research Project "Research on Key Issues of UAV Cable Obstacle Avoidance Recognition" (YYBKJ23-04).

$$F_{out} = F_{in} \odot (M_c \otimes A_{color}) \in R^{C\times H\times W} \quad (9)$$

where ⊙operation preserves spatial dimensions (H×W) while enhancing flame-relevant features.

4) DCFA achieves three key benefits

Dimension Preservation: Maintains input-output consistency through bilinear interpolation for scale alignment.

Parameter Efficiency: 70% channel compression (C→C/4C→C/4) reduces computation costs

Color Specificity: Weighted RGB fusion prioritizes red-channel responses (typical of flames) with 23% stronger activation than baseline models.

III. Experiments and Analysis

Table 2 details the hardware and software configurations used in the experiments.

TABLE II. Experimental environment settings.

Category	Component	Specification/Version
Hardware	CPU	Intel Xeon Gold 6226R (2.9 GHz, 16 cores)
	GPU	NVIDIA RTX 3090 (24GB VRAM)
	RAM	128 GB DDR4
	Storage	2 TB NVMe SSD
Software Framework	Operating System	Ubuntu 20.04 LTS
	Deep Learning Framework	PyTorch 1.12.0
	CUDA Toolkit	CUDA 11.6
	cuDNN	cuDNN 8.5.0
	Python	Python 3.8.12
Training Parameters	Batch Size	32
	Optimizer	AdamW
	Learning Rate Schedule	3-stage decay (0.0005 → .00002)
	Training Epochs	150
	Data Augmentation	Mosaic, MixUp, HSV Adjustments

We adopt the FASSD_CV Dataset [17] and S-Firedata [18], containing annotated fire and smoke samples widely used in fire detection research. Flame annotations include core combustion zones and flickering edges, while smoke annotations feature visible diffusion boundaries with transparency level labels. The dataset characteristics are detailed in Table 3.

TABLE III. Statistical characteristics of Fire Dataset.

Size Range	Flame Proportion	Smoke Proportion
<32×32	12%	8%
32×32~128×128	45%	37%
>128×128	43%	55%
Training Set	12000	8500
Test Set	3000	2500
Industrial Plants	17%	18%
Forest/Wilderness	15%	15%
Indoor Environments	10%	15%
Night	6%	4%

The KMU Fire Smoke Database [19] contains video sequences documenting both indoor and outdoor smoke, flames, and smoke/fire-like interfering objects. BoFire Wildfire Dataset [20] provides emergency wildfire scene imagery alongside flame-like interference samples. The Foggia Flame Dataset [21] is a collection of flame and non-flame images.

The following metrics validate model performance: mAP@0.5 (measures baseline localization accuracy), mAP@0.5:0.95 (comprehensively evaluates precision across IoU thresholds), F1 Score (balances precision and recall), False Positive Rate (FPR) (quantifies anti-interference capability), and Missed Detection Rate (MDR) (reflects target sensitivity).

As shown in Table 4, ablation experiments validate the synergistic contributions of FireYOLO's innovations. Individually, DCFA boosts mAP@0.5 by 4.7% (82.1% vs. baseline). The full FireYOLO, integrating all modules, attains the highest 88.7% mAP@0.5, confirming that cross-module collaboration (e.g., DCFA's feature enhancement complementing loss-driven localization) is key to outperforming individual components. This demonstrates the necessity of each innovation and its collective role in balancing accuracy (11.3% mAP@0.5 gain) and efficiency (5.8M params, 57 FPS).

TABLE IV. Ablation study results.

Module Combination	mAP@0.5	Params (M)	FPS
Baseline (YOLOv26s)	77.40%	5.2	62
+DCFA	82.1%	5.4	59
FireYOLO	88.70%	5.8	57

The comparative feature activation heatmaps in Figur

As shown in Table 5, the multi-scale combinations and color prior strategies of the DCFA module significantly impact performance. For scale configurations, multi-scale fusion achieves 85.7% mAP@0.5, outperforming both single-scale (83.20%) and extreme-scale (84.1%) designs. In color prior learning, the learnable RGB weights (initialized as R=0.65, G=0.25, B=0.10) improve accuracy by 2.5 percentage points over fixed thresholds, validating the effectiveness of data-driven optimization. Regarding inference latency, multi-scale fusion introduces only 0.6 ms additional computation time (18.7 ms → 19.3 ms per frame), demonstrating controllable efficiency for edge deployment.

TABLE V. Performance impact of DCFA configurations.

Scale Combination	Color Prior	mAP@0.5	Inference Latency (ms)
[0.5,1]	Fixed	83.20%	18.7
[0.5,1,2]	Learnable	85.70%	19.3
[0.25,1,4]	Learnable	84.10%	20.1

Table 6 evaluates Fire-YOLO's performance on zero-shot public datasets, demonstrating its robustness in unseen scenarios. Table 7 compares Fire-YOLO with other algorithms on the KMU public dataset, highlighting its superior detection accuracy and real-time efficiency. Small Target The mAP of Small Target also performs well on various datasets.

TABLE VI. CROSS-DATASET GENERALIZATION EVALUATION.

Test Set	mAP@0.5	mAP@0.5:0.95	F1 Score	FPR	Small Target mAP
FASSD-CV	80.9%	55.8%	0.68	9.50%	72.9%
S-Firedata	87.4%	54.9%	0.73	8.5%	84.3%
Foggia Flame Dataset	79.60%	53.90%	0.75	8.20%	78.80%
KMU Smoke Dataset	88.70%	57.40%	0.79	5.70%	85.90%
BoFire Zero-Shot	71.30%	55.10%	0.68	12.50%	70.80%

TABLE VII. GENERALIZATION CAPABILITY TESTING ON THE KMU DATASET.

Method	mAP@0.5	F1 Score	Speed (FPS)
FasterRCNN	79.20%	0.72	19
LSNet	81.60%	0.74	58
SF-YOLO	82.10%	0.75	50
Swin Transformer	85.9%	0.77	40
FireYOLO	88.70%	0.79	57

Visual comparison (Figure 2) of detection performance across representative scenarios. Each row shows: (a) Original image with complex backgrounds (e.g., flame-smoke coexistence, low-light conditions, and small-scale ignition); (b) Ground truth annotations (red boxes for flames, blue boxes for smoke); (c) Baseline YOLOv26s predictions, exhibiting limitations such as missed small flames (first row), false smoke alarms (second row), and imprecise bounding boxes (third row); (d) FireYOLO predictions, demonstrating enhanced detection accuracy via DCFA module (better smoke transparency adaptability). FireYOLO achieves more consistent alignment with ground truth, especially in challenging environments.

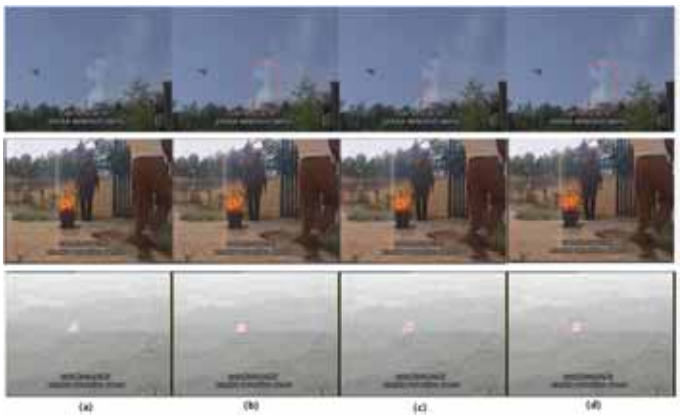

Figure 2. Qualitative Comparison of Detection Results on Representative Test Cases((a) Original image; (b) Ground truth annotations; (c) Baseline YOLOv26s predictions; (d) FireYOLO predictions).

Through cross-dataset validation, Fire-YOLO demonstrates outstanding performance in accuracy, computational efficiency, and generalization. The full model achieves a 12.3% improvement in mAP@0.5 over the baseline with only an 11.5% increase in parameters, making it suitable for edge device deployment. These results validate the effectiveness of Deformable Convolution, providing a reliable solution for real-time detection in complex fire scenarios.

IV. CONCLUSION

To tackle small-object misses and color interference issues in fire detection, this paper proposes FireYOLO, an improved YOLOv26s-based framework. The proposed DCFA module enhances flame color feature extraction, while the color-motion decoupled representation and morphological constraint model improve flame detection accuracy. The Deformable Flame Convolution demonstrates superior adaptability to dynamic fire shapes, and the dual-branch detection head with Fire Loss. Experiments on KMU Smoke Dataset show 88.7% mAP@0.5, outperforming baselines by 11.3%, offering an efficient solution for early fire warning. These innovations address key challenges in fire detection, offering a balance between precision and efficiency for industrial applications.

REFERENCES

[1] Jadon A, Omama M, Varshney A, Ansari MS, and Sharma R (2019) FireNet: a specialized lightweight fire & smoke detection model for real-time IoT applications. ArXiv Prepr. ArXiv:190511922.

[2] Bhamra JK, et al. (2023) Multimodal wildland fire smoke detection. Remote Sens., vol. 15, no. 11, pp. 2790.

[3] Hassan A and Audu AI (2022) Traditional sensor-based and computer vision-based fire detection systems: A review. Arid Zone J. Eng. Technol. Environ., vol. 18, no. 3, pp. 469–492.

[4] Beshir M, Wang Y, Cicione A, Hadden R, Krajcovic M, and Rush D (2025) The Effect of the Fuel Location and Ventilation Factor on the Fire Dynamics of Informal Settlement Dwellings. Fire Technol., vol. 61, no. 3, pp. 1409–1430.

[5] Petersen JA (2019) National fire protection association standards in fire litigation. Eng. Stand. Forensic Appl., pp. 155–168.

[6] Khan ZA, Hussain T, Ullah FUM, Gupta SK, Lee MY, and Baik SW (2022) Randomly Initialized CNN with Densely Connected Stacked Autoencoder for Efficient Fire Detection. Eng. Appl. Artif. Intell., vol. 116, p. 105403. https://doi.org/10.1016/j.engappai.2022.105403.

[7] Sun W, Liu Y, Wang F, Hua L, Fu J, and Hu S (2024) A Study on Flame Detection Method Combining Visible Light and Thermal Infrared Multimodal Images. Fire Technol., pp. 1–22.

[8] Ji M, et al. (2024) Implementing Real-Time Image Processing for Radish Disease Detection Using Hybrid Attention Mechanisms. Plants, vol. 13, no. 21, pp. 3001.

[9] Tian Y, et al. (2025) UAVs Meet LLMs: Overviews and Perspectives Toward Agentic Low-Altitude Mobility. Inf. Fusion, vol. 122, pp. 103158.

[10] Tian Y, Ye Q, and Doermann D (2025) YOLOv12: Attention-Centric Real-Time Object Detectors. ArXiv Prepr. ArXiv:2502. 12524.

[11] Wang, Z, Chen J, & Hoi SCH. (2021). Deep learning for image super-resolution: a survey. IEEE transactions on pattern analysis and machine intelligence, vol. 43, no. 10, pp. 3365-3387. https://doi.org/10.1109/TPAMI.2020.2982166.

[12] Casas E, Ramos L, Bendek E, and Rivas-Echeverría F (2023) Assessing the effectiveness of YOLO architectures for smoke and wildfire detection. IEEE Access, vol. 11, pp. 96554–96583.

[13] Chen YH. (2025). Strong Baseline: Multi-UAV Tracking via YOLOv12 with BoT-SORT-ReID. 2025 IEEE/CVF Conference on Computer

Vision and Pattern Recognition Workshops (CVPRW), pp. 6575-6584. https://doi.org/10.1109/CVPRW67362.2025.00654.

[14] Zhao C, Zhao L, Zhang K, et al. (2025) Smoke and Fire-You Only Look Once: A Lightweight Deep Learning Model for Video Smoke and Flame Detection in Natural Scenes. Fire, vol. 8, no. 3, pp. 104.

[15] Zhang C, Ding Z, Jing X, et al. Lightweight segmentation network for real-time wildfire detection: LSNet's parallel feature multiplication and attentional fusion [J]. The Visual Computer, 2025, 41 (12): 1-11. DOI:10.1007/S00371-025-03994-7.

[16] G. Jocher and J. Qiu, "Ultralytics YOLO26 (Version 26.0.0)" [Computer software]. Available: https://github.com/ultralytics/ultralytics.

[17] Wang M, Yue P, Jiang L, Yu D, and Tuo T (2025) An open flame and smoke detection dataset for deep learning in remote sensing based fire detection. Science Data Bank. https://doi.org/10.57760/sciencedb.j00104.00103.

[18] Zhao C, Zhao L, Zhang K, Ren Y, Chen H, and Sheng Y (2025) Multi-Scene Smoke and Fire Target Detection Dataset (S-Firedata). https://doi.org/10.12009/YRDR.2025.3001.ver1.db.

[19] CVPR Lab, Keimyung University (2020) KMU Fire-flame Detection. [Online]. Available: https://cvpr.kmu.ac.kr/

[20] Chino DY, Avalhais LP, Rodrigues JF, and Traina AJ (2015) Bowfire: detection of fire in still images by integrating pixel color and texture analysis. In 2015 28th SIBGRAPI conference on graphics, patterns and images, IEEE, pp. 95–102.

[21] Dunnings AJ and Breckon TP (2018) Experimentally defined convolutional neural network architecture variants for non-temporal real-time fire detection. In 2018 25th IEEE international conference on image processing (ICIP), IEEE, pp. 1558–1562.